As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 8(06): 246-254, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0806043
Available online at: http://www.ijmtst.com/vol8issue06.html

Parkinson's Disease Detection Using Biomedical Voice Measurements

Dr.Preety, Gaurav Kumar, Akshaya Renganathan , Pooja, Rahul Kamra

Department of Information Technology, Dr Akhilesh Das Gupta Institute of Technology and Management, India.
Corresponding Author Email Id: laxaman.pal@gmail.com

To Cite this Article
Dr.Preety, Gaurav Kumar, Akshaya Renganathan , Pooja and Rahul Kamra. Parkinson's Disease Detection Using Biomedical Voice Measurements. International Journal for Modern Trends in Science and Technology 2022, 8(06), pp. 246-254. https://doi.org/10.46501/IJMTST0806043

Article Info
Received: 12 May 2022; Accepted: 06 June 2022; Published: 10 June 2022.

ABSTRACT

The Parkinson's disease is moderate neuro degenerative confusion that influences a ton just individuals fundamentally influencing their personal satisfaction. It for the most part influence the engine elements of human. The really engine side effects are designated "parkinsonism" or "parkinsonian disorder". The side effects of Parkinson's disease will happen gradually, the side effects incorporate shaking, inflexibility, gradualness of development and trouble with strolling, Thinking and conduct change, Depression and uneasiness are additionally normal. There is a model for identifying Parkinson's utilizing voice. The avoidances in the voice will affirm the side effects of Parkinson's disease. This undertaking showed 73.8% productivity. In our model, a gigantic measure of information is gathered from the typical individual and furthermore recently impacted individual by Parkinson's disease. This information is prepared utilizing AI calculations. From the entire information 60% is utilized for preparing and 40% is utilized for testing. The information of any individual can be placed in database to check regardless of whether the individual is impacted by Parkinsons disease. There are 24 sections in the informational index every segment will show the side effect upsides of a patient with the exception of the status segment. The status section has 0's and I's, those values will conclude the individual is affected with Parkinsons disease. I's show individual is affected, 0's demonstrate typical circumstances.

KEYWORDS: Parkinson's disease; machine learning (ML), XG Boost, Random Forest , SVM , KNN

1. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative clinical disorder characterized by dopamine receptor dysfunction[7]. The lack of movement seems to be the most prominent indication of Parkinson's disease. As a result, a person's mobility may become quite stagnant. Parkinson's dementia is a progressive illness that manifests itself in both motor and non-motor symptoms[11]. Aside from numerous similar symptoms, each person will experience and display a unique presentation of the ailment. Clinical symptoms of Parkinsonism include erratic power (lack of movement), bradykinesia (lack of control of movement), stiffness (wrist, shoulder, and neck), and rest tremors, are frequent in patients who have the disease (I mbalance of neurotransmitters, dopamine and acetylcholine)[17],[18]. Parkinsonism can also be caused by medications, as well as less familiar diseases like multiple cerebral infarction and degenerative disorders like progressive supra nuclear palsy (PSP) and multiple system atrophy (MSA).

Parkinson's disease patients appear rigid and inflexible. Suffering from Parkinson's disease patients may look "frozen" or unable to move at times. Symptoms and complications of Parkinson's disease, the depletion of dopamine-producing nerve cells induces a neurological condition.Machine learning (ML) techniques have proved the capacity to handle enormous amounts of medical datasets and have provided perceptual guidance. ML-based solutions may increase individual safety, enhance clinical care quality, reduce healthcare costs, and aid clinicians' efforts by processing large data sets from patient information[19].ML techniques have been widely used for disease categorization and prediction. Machine learning algorithms use unsupervised algorithms to diminish the complexity of the data, allowing the disorder to be diagnosed. These strategies also enable for data manipulation, data de - noising, similarity calculations, and data segmentation.On the other hand, supervised learning algorithms are utilized to enable the disease's ultimate categorization, prediction, and diagnosis. While ML has shown to be beneficial, its proper deployment necessitates a significant effort on the part of human professionals, given that no one technique can produce acceptable results in all potential cases.

To determine which algorithm is most suited for detecting illness beginning, we will employ XGBoost, KNN, SVMs, and the Random Forest Algorithm.

OBJECTIVE:

A series of research out of discovered that there was a linkage between Parkinson's disease and speech defects, and that most persons with Parkinson's disease had speech problems and difficulty in the early stages. As an outcome, voice attributes will play a prominent part in early disease detection. Because it is reliant on medical professionals' experience, diagnosing PD, like other neurological diseases, is exceedingly difficult and expensive. Because of the difficulty in order to diagnose and treat coronary artery disease requires the patient to visit the hospital thoroughly every time, elderly patients are severely affected because For them, it isquite inconvenient. It's critical to have a viable early diagnostic and treatment model that medical professionals can employ, especially in places where medical experts are few. The major goal of this study was to investigate if these speech traits may be utilised to detect Parkinson's disease early on.

PROBLEM STATEMENT:

Diagnosis of Parkinson's Disease usually requires a patient's record of sensitivity in terms of muscle skills in most cases. It becomes more difficult for a doctor in the early stages during a diagnosis when the patient's results are not so bad. The patient needs to visit the clinic regularly to monitor the progression of the disease over time. A productive diagnostic procedure does not require a visit from a doctor and can be very helpful. People with Parkinson's Disease show different voice symptoms, so voice recording is considered a useful diagnostic tool. The use of machine learning algorithms in the speech database to diagnose accurate disease can be a productive diagnostic step before visiting adoctor.

2. LITERATURESURVEY

Several publications on the diagnosis of Parkinson's disease using verbal data have been published in recent years. The author developed a model for PD detection that included Neural Networks, Decision Trees, Backbones, and DMneural, as well as a comparison analysis[2]. For MFCC voice recording samples, a paper employed the (LOSO) verification technique and the SVM phase for screening between PD patients and healthy individuals[3]. Ali suggested a two-dimensional sample and feature selection method for PD detection as a preliminary prediction model.[1].Another paper describes the selection of two-dimensional attributes and a separation model for the detection of PD [4]. One author has used a parallel reading strategy that combines different categories to obtain PD with an accuracy of 86% [6]. Other articles have proposed models for PD detection that have a good level of accuracy. For instance, a composite paper with a combined weight and a proposed model based on the Complex Valued Artificial Neural Network (CVANN) achieved 99.5 percent accuracy. Despite its excellent accuracy, however, this test produces biassed outcomes. The experiment used small data points in the provided data sets, and each individual had several audio recording data. [5].

Sonu et al. [21] applied computational tree classification and recurrence (CART) in predicting PD using a voice recording dataset from the UCI store.

Aich.et.al. [22] analyzed seven different AI approaches using two types of element identification methods (PCA and genomic computation-based inclusion sets).They also found that the GA-based bright spot determination beat the SVM when they used a different measure of the positive pre-prefix, and continued to overflow other classifiers using the solution.

Prashanth et al. [23] used an estimate of CSF of preclinical markers of non-motor highlights of RBD and olfactory misfortune, and imaging highlights of dopaminergic to organize early identification of PDs and healthy normal individuals to extend the work done by another creator. They achieved the highest accuracy of 96.4% with the SVM.

While L.Parisi et al. [24] analyzed their work on PD ordering using MLP computation LSVM with different surveys and affordable programming using three validations different element identifiers. They found that their MLPLSVM calculation gave the best accuracy, being a computationally powerful and fast classifier in recognizing PDs from speech information.

Mr.Nilashi et al. [25] tried to advocate a mixed intelligence technique with a score of to reduce computation time with the desire to work on the accuracy of their predictions. they incorporated-dimensional reduction and clustering into their process. They achieved an accuracy of 86.8%. applied unique AI calculations in PD prediction using the

speech quantity dataset. Their performed test consisted of a selection with Pearson's correlation scoring strategy and use 10 out of 22 entries in their survey. They achieved

95% accuracy with Augmented Decision Trees, which is an illustration of the

hold process using a drop tilt that helps the classifier quickly give accurate results.

Agarwal et al. [26] used Extreme Learning Machine (ELM) to predict PD using voice signals. as the best result obtained by ELM achieved 90.76 percent accuracy with 500 neurons and active sigmoid sections. Using principal component analysis (PCA), they were able to minimize the size of the highlights to just 13. When looking for linear recurrence and support vector recurrence, they discovered that their method generated better results. The model containing nine highlights from the first 18 factors had the highest accuracy of 85.60 percent.

While J. Hughes et.al.[27] analyzed ANN and XGBoost models at Symbolic Regression (SR)based on previous work using a patient gait dataset (this is a side effect of PDs considering the gait progression). They consider this as even though every calculation matches the sign, XGBoost also ANN gives the best result. In all cases, the XGBoost calculation matches the availability information better and is able to summarize up to hidden information better than the ANN. Other visualizers. advancement and viability, the XGBoost has all the hallmarks of being a promising methodology for the future.

Rahman et al. [28] applied four algorithms in the prediction of PD by removing both standard acoustic factors (Mel frequency coefficient (MFCC), jitter and luminous variation). and algorithms in light of using voice information. They came up with a display of 0.75 AUC (Area Under the Curve) after displaying the standard audio highlights. via XGBoost. it is fast and uses modern heuristics to punish inefficient trees and makes better use of regularization. Database Information Discovery (KDD) workflow was used during project implementation [29].

Despite having several published papers on the discovery of PD, improvements are still needed to provide more accurate, robust and effective models.

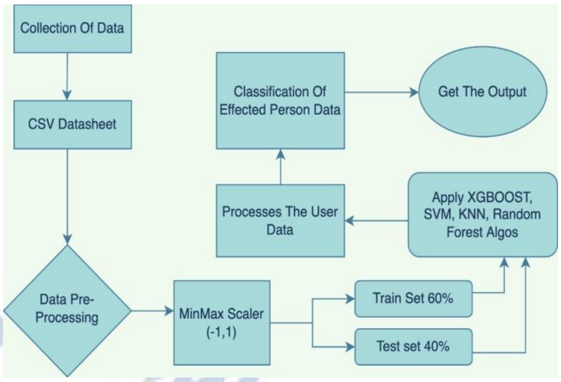
3.DATA PREPARATION

This dataset is made up of 31 biomedical voice measures, 23 of whom have Parkinson's disease (PD)[8]. Each column in the table refers to a certain voice measure, and each row corresponds to one of the 195 voice recordings made by these people ("name" column)[9. The primary goal of the data is to distinguish healthy persons from those with Parkinson's disease using the "status" column, which is set to 0 for healthy people and 1 for those with PD. The information is stored in ASCII CSV format. Each row in the CSV file corresponds to a single voice recording instance. There are around six recordings per patient, with the patient's name in the first column. A training dataset is a collection of data that is used to train a model. The testing dataset is used to evaluate the trained model's performance.

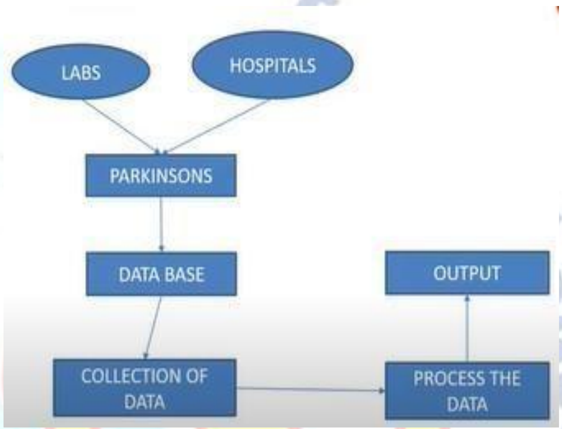
Table I: Features used from Dataset

S No.	Name	ASCII subject
1.	MDVP:Fo(Hz)	Averagevocal fundamental frequency
2.	MDVP:Fhi(Hz)	Maximum vocal fundamental frequency
3.	MDVP:Flo(Hz)	Minimum vocal fundamental frequency
4.	MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP	Several measures of variation in fundamental frequency
5.	MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA	Several measures of variation in amplitude
6.	NHR,HNR	Two measures of ratio of noise to tonal components in the voice
7.	status	Health status of the subject (one) - Parkinson's, (zero) - healthy
8.	RPDE,D2	Two nonlinear dynamical complexity measures
9.	DFA	Signal fractal scalingexponent
10.	spread1,spread2,PPE	Three nonlinear measures of fundamental frequencyvariation

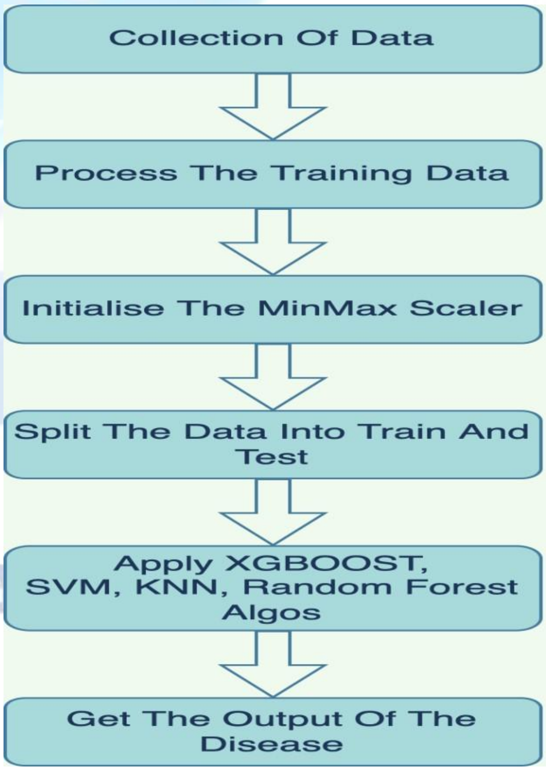
Architecture Diagram:



ER Diagram:



Data Flow Diagram:



4. DESIGN AND METHODOLOGIES

The different algorithms explored in this paper are listed as below:

XGBoost: XGBoost is a Gradient Boosted decision tree implementation. This approach generates decision trees sequentially. Weights are crucial in XGBoost[11]. Weights are assigned to each independent variable, which are then input into the decision tree, which predicts results. The strength of variables predicted incorrectly by the tree is increased, and these variables are subsequently put into the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined predictionproblems.

SVM: The Support Vector Machine, or SVM, is a popular Supervised Learning technique that may be used to solve both classification and regression issues. However, it is mostly utilised in Machine Learning for classification difficulties.[15].

The SVM algorithm's goal is to find the best line or decision boundary for categorising n-dimensional space so that we can easily place new data points in the correct category in the future. A hyperplane is the best decision boundary.

SVM selects the extreme points/vectors that aid in the creation of the hyperplane. These extreme cases are referred to as support vectors, and the algorithm is known as the Support Vector Machine.

KNN: The KNN algorithm is a sort of supervised machine learning method that may be used to solve both classification and regression predicting problems. The KNN algorithm predicts the values of new datapoints based on 'feature similarity,' which implies that the new data point will be assigned a value depending on how closely it matches the points in the training set. [20] The KNN algorithm is a sort of supervised machine learning method that may be used to solve both classification and regression predicting problems.

The KNN algorithm predicts the values of new datapoints based on 'feature similarity,' which implies that the new data point will be assigned a value depending on how closely it matches the points in the training set.

K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

Random Forest: Random Forest algorithms are utilised for both bracketing and retrogression. It generates a tree for the data and bases vaticination on it. When significant sets of record values are missing, the Random Forest technique can still provide the same results. The decision tree samples that are created can be stored and used on different data[10]. In arbitrary timber there are two stages, originally produce an arbitrary timber also make a vaticination using an arbitrary timber classifier created in the first stage.

Methodology :

Data Collection: Data processing is a key step in any machine learning problem. We will use a database from Kaggle for this issue. This database is based on a list of [8]biomedical voice measurements from 31, 23 people with Parkinson's disease (PD). Each column in the table is a measure of a specific word, and each row corresponds to one of the 195 voice recordings for these people ("word" column)[9]. The main purpose of the data is to discriminate against healthy people and those who have PD, according to the "status" column set to 0 healthy and 1 in PD. Data is in ASCII CSV format. CSV file lines contain an event associated with a single voice recording. There are approximately six recordings for each patient, the patient's name is identified in the first column.

Training and testing of data: Cleaning is a very important step in a machine learning project. The quality of our data determines the quality of our machine learning model. It is therefore always necessary to clean the data before feeding it to the training model. In our database all numerical columns, the target column i.e. prediction is a type of character unit and is coded using a numeric codec[14].

Code completion: After collecting and cleaning data, the data is ready and can be used to train the machine learning model. We will use here XGBoost, KNN Algorithm, Support Vector Machines (SVMs), Random Forest Algorithm and use the data set found in the UCL Parkinson data set. We will now be predicting the disease with the inclusion and accuracy of each algorithm. This makes our whole assumptions stronger and more accurate. Finally, we will define a function that captures comma-separated symptoms as input, predicts disease based on symptoms using trained models, and determines whether a person has the disease or not. Then find the accuracy of all the algorithms and models and then again train the model and compare the accuracy and predict the best model with the highest accuracy.

5. IMPLEMENTATION

- First dataset is uploaded and run for theprediction.

Reading the file data of Parkinson disease

```
df=pd.read_csv('/content/parkinsons.csv')
df.head()
```

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDV
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	

5 rows × 24 columns

- Now train the model and test it to get the accuracy and output for datasetinput.

- Now let start one by one with our algorithms

XGBoost:

```
Accuracy : 94.87 %
Mean Absolute Error: 5.13 %
```

	Predicted Healthy	Predicted Parkinsons
True Healthy	5	2
True Parkinsons	0	32

```
119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.0
Parkinson's Disease Detected
```

SVM:

```
Accuracy : 87.18 %
Mean Absolute Error: 12.82 %
```

	Predicted Healthy	Predicted Parkinsons
True Healthy	2	5
True Parkinsons	0	32

```
119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.0
Parkinson's Disease Detected
```

KNN:

```
Accuracy : 89.74 %
Mean Absolute Error: 10.26 %
```

	Predicted Healthy	Predicted Parkinsons
True Healthy	4	3
True Parkinsons	1	31

```
119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.0
Parkinson's Disease Detected
```

Random Forest:

```
Accuracy : 92.31 %
Mean Absolute Error: 7.6899999999999995 %
```

	Predicted Healthy	Predicted Parkinsons
True Healthy	4	3
True Parkinsons	1	31

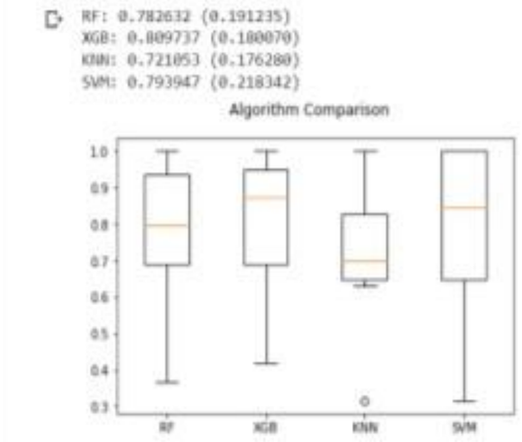
```
119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.0
Parkinson's Disease Detected
```

- Now again train the model and compare their accuracies atonce
- Comparison of theiraccuracies:

	MLA used	Train Accuracy	Test Accuracy	Precission	Recall	AUC
1	XGBClassifier	100.0000	93.8776	0.925000	1.000000	0.875000
2	SVC	81.5068	83.6735	0.822222	1.000000	0.666667
3	KNeighborsClassifier	90.4110	83.6735	0.871795	0.918919	0.751126
4	RandomForestClassifier	100.0000	91.8367	0.923077	0.972973	0.861486

- Taking the mean standard deviation of the algorithms we arecomparing

Box-plot Algorithm Analysis



6. RESULT ANDANALYSIS

This section displays the results obtained by using XGBoost, SVM, KNN, and Random Forest. Delicacy score, Precision (P), Recall (R), and AUC are the criteria used to assess the algorithm's performance. The precision metric (given in equation (2)) offers a correct measure of positive analysis. The measure of correct factual cons (stated in equation (3)) is called recall. Area Under the Curve (AUC) is calculated using the trapezoidal rule.

- Precision = (TP +FP) / (TP) (2)

(TP) / (TP+FN) = Recall (3)

- F– Measuring =(2 * Precision + Recall) / (Precision + Recall) (4)
- TP True positive means that the case has a complaint and the test results are positive.
- FP False positive: The case does not contain a complaint, but the test results are positive.
- TN True Negative: There is no complaint in the case, and the test results are negative.
- FN False negative: Although the case involves a complaint, the test results are negative.

In the trial, the pre-processed dataset is used to run the tests, and the techniques indicated above are studied and applied. The confusion matrix is used to achieve the performance criteria listed below. The model's performance is described by the Confusion Matrix.. The confusion matrix attained by the proposed model for different algorithms is shown below in Table 2.The delicacy score Attained for XGBoost , SVM, KNN and Random Forest is shown below inTable3.

TABLE II.VALUES OBTAINED FOR CONFUSION MATRIX USING DIFFERENT ALGORITHM

Algorithm	True Positive	False Positive	False Negative	True Negative
XGBoost	5	0	2	32
SVM	2	0	5	32
KNN	4	1	3	31
Random Forest	11	2	1	35

TABLE III. ANALYSIS OF MACHINE LEARNING ALGORITHM

Algorithm	Precision	Recall	AUC	Accuracy
XGBoost	0.925	1.00	0.875	94.87%
SVM	0.822	1.00	0.666	87.17%
KNN	0.871	0.918	0.751	89.74%
Random Forest	0.900	0.972	0.819	93.88%

7. CONCLUSION

This work is an attempt to present a wide assessment of Parkinson disease detection systems that have used various machine learning approaches. The overview of results acquired by various researchers is made accessible in the literature review section table; practically all authors/researchers made significant attempts to forecast Parkinson disease using unique methodologies. The majority of all ML strategies used by various authors performed better, although constructing a highly fast classifier employing unique neural network architecture combined with a specific strategy may perform better. The study's goal was to create the most effective machine learning algorithm for Parkinson's disease discovery. Using the UCI machine literacy depository dataset, this study analyses the accuracy scores of the XGBoost, SVM, KNN, and Random Forest algorithms for Parkinson's disease. The results of this study show that the XGBoost algorithm has a higher accuracy than others, at 94.87 percent, and that it is based on the XGBoost algorithm as well as using a larger dataset than the one used in this analysis, which will help to produce better results and assist health professionals in effectively and efficiently prognosticating parkinson's disease.

8. FUTURE SCOPE

The scope of the challenge right here is that integrating medical choice assist with automatic affected person information can lessen clinical errors, enhance affected person safety, lessen undesirable modifications in practice and enhance affected person outcomes. This thought is quiet promising because of fact statistics modeling and evaluation gear have the capability to create a knowledge-wealthy surroundings which could enhance the great of medical selections.

Effective implementation of device studying permits healthcare experts to make higher selections, perceive developments and innovations, and enhance the performance of studies and medical trials.The scope of the project here is that combining clinical decision support with computerized patient records can reduce medical errors, improve patient safety, reduce unnecessary changes in improve patient outcomes and performance . This proposal is promising because the tools for modeling and data analysis have the potential to create an enriching environment that can greatly improve the quality of clinical decisions.

Effective implementation of machine learning enables health care professionals to make better decisions, identify trends and innovations, and improve the effectiveness of clinical research and trials.

Through machine learning, physicians Healthcare providers can make better decisions about their patients' diagnoses and treatment options, leading to the full development of medical field. Machine learning algorithms can be helpful in providing critical statistics, real-time data, and advanced patient disease analysis, which are processed on computer devices that can take into account a large number of variables, this means more accurate health data. According to a recent study, researchers found better diagnostic accuracy, using complete medical records looking at about 200 variables,

With device studying, clinicians Healthcare companies could make higher selections approximately their patients' prognosis and remedy options, main to an standard development in fitness services. Machine studying algorithms also can be beneficial in offering essential statistics, real-time statistics, and superior evaluation of a affected person's illness, processed on computing gadgets which could contemplating a massive quantity of variables, this indicates greater correct fitness statistics. According to a current study, researchers received higher diagnostic accuracy, the usage of entire clinical information searching at approximately 2 hundred variables.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Ali, L., Zhu, C., Zhou, M., Liu, Y.:(2019) Early diagnosis of parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. Expert Systems with Applications 137, 22–28

[2] Benba, A., Jilbab, A., Hammouch, A(2016).: Voice assessments for detecting patientswith parkinson's diseases using pca and npca. International Journal of SpeechTechnology 19(4), 743–754.

[3] Cant¨urk,I., Karabiber, F(2016).Arabian Journal for Science and Engineering :A machine learning system for the diagnosis of parkinson's disease from speech signals and its application to multiple speech signal types. 41(12), 5049–5059.

[4] Naranjo, L., P´erez, C.J., Campos-Roca, Y., Mart´ın, J(2016).: Addressing voice recordingreplications for parkinson's disease detection. Expert Systems with Applications46, 286–292.

[5] Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin ,H., Kursun, O(2013)IEEE Journal of Biomedical and Health Informatics: Collection and analysis of a parkinson speech dataset with multipletypes of sound recordings17(4), 828–834.

[6] Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu,M., Aydin, T., Isenkul, M.E., Apaydin, H. (2019): A comparative analysis of speech sig-nal processing algorithms for parkinson's disease classification and the use of thetunable q-factor wavelet transform. Applied Soft Computing 74, 255–263.

[7] Jankovic, J. (2008) Journal of neu-rology, neurosurgery & psychiatry: Parkinson's disease: clinical features and diagnosis. 79(4), 368–376.

[8] UCI Machine Learning Repository, 'Parkinson Data Set'. [Online]. Center for Machine Learning and Intelligent Systems Available: [Accessed on: Aug. 12, 2021]. https://archive.ics.uci.edu/ml/datasets/parkinsons

[9] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, 'Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease', IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp. 1015–1022, Apr. 2009, doi: 10.1109/TBME.2008.2005954. [Accessed on: Aug. 12, 2021].

[10] A. Liaw, M. Wiener et al. (2002)., "Classification and regression by randomforest," R news, vol. 2, no. 3, pp. 18–22,

[11] Z. Xuchen, F. Yong, and W. Peng (2019), 'Automatically Predicting Severity of Parkinson's Disease Using Model Based on XGBoost from Speech', IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Dalian, China, Sep. 2019, pp. 1–5. doi: 10.1109/ICSPCC46631.2019.8960722. Available at: https://ieeexplore.ieee.org/document/8960722 [Accessed on: Aug. 12, 2021].

[12] Parkinson's Foundation 2021, "What is Parkinson's?". [Online].Available at: https://www.parkinson.org/understanding-parkinsons/what-is-parkinsons [Accessed on: Aug. 12, 2021]. 17(4), 828–834 2013

[13] S. Aich, H.-C. Kim, K. younga, K. L. Hui, A. A. Al-Absi, and M. Sain(2019) 21st International Conference on Advanced Communication Technology (ICACT, 'A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease',), PyeongChang Kwangwoon_Do, Korea (South), Feb. 2019, pp. 1116–1121. doi: 10.23919/ICACT.2019.8701961. Available at: A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease | IEEE Conference Publication | IEEE Xplore [Accessed on: Aug. 12, 2021].

[14] P. Hart.(1968): The Condensed Nearest Neighbour Rule, IEEE Transactions on Information Theory, 14, 515-516,

[15] Akbani, R.; Kwek1, S.; Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets, in Proceedings of European Conference on Machine Learning, Lecture Notes in Computer Science 3201:39–50

[16] Lorena Rosa S Almeida,Guilherme Valenca,Jamary Oliveira-Filho(2017).Journal of the Neurological Sciences:A MDS-UPDRS Part II (motor experiences of daily living) score to predict recurrent falls in people with Parkinson's disease 383:114-115 DOI:10.1016/j.jns.2017.10.047

[17] D. Heisters(2011),British Journal of Nursing, "Parkinson's: symptoms, treatments and research,", vol. 20, no. 9, pp. 548–554,

[18] J. Farlow, N. D. Pankratz, J. Wojcieszek, and T. Foroud(2014), "Parkinsondisease overview,"

[19] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh(2016) ,International Journal of Medical Informatics, 'High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning' , vol. 90, pp. 13–21, Jun. , doi: 10.1016/j.ijmedinf.2016.03.001. Available at: High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning - ScienceDirect [Accessed on: Aug. 12, 2021].

[20] P. Hart.(1968): The Condensed Nearest Neighbour Rule, IEEE Transactions on Information Theory, 14, 515-516,

[21] S. R. Sonu, V. Prakash, R. Ranjan, and K. Saritha (Aug. 2017), 'Prediction of Parkinson's disease using data mining, in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, pp. 1082–1085. doi:10.1109/ICECDS.2017.8389605. Available at: Prediction of Parkinson's disease using data mining | IEEE Conference Publication | IEEE Xplore [Accessed on: Aug. 12, 2021].

[22] S. Aich, H.-C. Kim, K.younga, K. L. Hui, A. A. Al-Absi, and M. Sain(2019), 'A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease', 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), Feb. 2019, pp. 1116–1121. doi: 10.23919/ICACT.2019.8701961. Available at: A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease | IEEE Conference Publication | IEEE Xplore [Accessed on: Aug. 12, 2021].

[23] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh(Jun. 2016), 'High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning, International Journal of Medical Informatics, vol. 90, pp. 13–21, , doi: 10.1016/j.ijmedinf.2016.03.001. Available at: High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning - ScienceDirect [Accessed on: Aug. 12, 2021].

[24] L. Parisi, N. Ravichandran, and M. L. Manaog,(Nov. 2018), 'Feature-driven machine learning to improve early diagnosis of Parkinson's disease, Expert Systems with Applications, vol. 110, pp. 182–190, doi: 10.1016/j.eswa.2018.06.003. Available at: Feature-driven machine learning to improve early diagnosis of Parkinson's disease - ScienceDirect [Accessed on: Aug. 12, 2021].

[25] M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, and M. Farahmand(2018), 'A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques', Biocybernetics and Biomedical Engineering, vol. 38, no. 1, pp. 1–15, doi: 10.1016/j.bbe.2017.09.002. Available at: A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques - ScienceDirect [Accessed on: Aug. 12, 2021].

[26] A. Agarwal, S. Chandrayan, and S. S. Sahu(Mar. 2016), 'Prediction of Parkinson's disease using speech signal with Extreme Learning Machine', in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, pp. 3776–3779. doi: 10.1109/ICEEOT.2016.7755419. Available at: Prediction of Parkinson's disease using speech signal with Extreme Learning Machine | IEEE Conference Publication | IEEE Xplore [Accessed on: Aug. 12, 2021].

[27] J. A. Hughes, S. Houghten, and J. A. Brown(Nov. 2020), 'Models of Parkinson's Disease Patient Gait', IEEE J. Biomed. Health Inform., vol. 24, no. 11, pp. 3103–3110, doi: 10.1109/JBHI.2019.2961808. Available at: https://ieeexplore.ieee.org/abstract/document/8939380 [Accessed on: Aug. 12, 2021].

[28] W. Rahman et al.(Dec. 2020), 'Detecting Parkinson's Disease From an Online Speech-task', arXiv:2009.01231 [cs, eess, stat], Accessed: Jul. 03, 2021. [Online]. Available: http://arxiv.org/abs/2009.01231

[29] G . Fayyad, Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy(1996), "From Data Mining to Knowledge Discovery: An Overview", Advances in Knowledge Discovery and Data Mining, AAAI Press / The MIT Press, Menlo Park, CA, pp.1-34 Available at: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html [Accessed on: Aug. 12, 2021].